NLSY97 Appendix 4:

Geographic Variable Creation

Several variables in the main data set provide information about the respondent's area of residence. These variables permit researchers to identify key characteristics of the area without needing access to the geocode CD-ROM. Geographic variables were created using a software program called Maptitude; therefore, no programming code is provided for these variables. Instead, this document offers a brief general description of the methods used to generate these variables. For more information about the process of classifying a respondent's metropolitan area or about the geographic variables in general, refer to the introduction to the *Geocode Codebook Supplement* or contact NLS User Services.

> ***User Notes:*** Researchers should be aware that the process for geocoding respondent residences has been changed since the initial survey round. All residence information was re-geocoded using the new, more accurate approach, so all variables are comparable across rounds. However, researchers should not use geographic variables from old CDs in analyses; all geographic data should be taken from the newest CD.

CENSUS REGION OF RESIDENCE AT SURVEY DATE

Variable Created: CV_CENSUS_REGION

This variable classifies respondents as residing in one of four regions defined by the U.S. Census Bureau. These regions are as follows:

Census Division	States
Northeast	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont
North Central	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin
South	Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia
West	Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming

MSA STATUS AT SURVEY DATE

Variable Created: CV_MSA

This variable provides users with information about whether the respondent lived in the central city of the MSA, in another part of the MSA, or outside of an MSA. As defined by the Census Bureau, a central city is the major city lying within a Metropolitan Statistical Area (MSA). Initially, a variable was created using the TIGER/Line files (a database developed by the Census Bureau; the round 4 variable used the 1998 version) to determine whether the respondent lived in an MSA. A second variable was created based on "places" data in the Maptitude program that identified whether the respondent lived in the central city. The version of Maptitude used for round 4 used the 1992 TIGER/Line files to classify places as central cities. The variables were then combined to produce a single MSA/central city variable. For this dataset, respondents are coded as follows:

1 not in MSA
2 in MSA, not central city
3 in MSA, central city
4 in MSA, not known
5 not in country

RURAL VS. URBAN

Variable Created: CV_URBAN_RURAL

Places are identified as urban or rural by the Census Bureau. Urban places are those in "urbanized areas" or "places" with a population of at least 2,500; all other areas are rural. According to the Census Bureau, about 25 percent of the U.S. population lives in rural areas. The "urbanized area" map in the Maptitude software used in round 4 was based on the 1990 Census Bureau TIGER/Line files. Respondents residing in urbanized areas are coded 1 and those residing in rural areas are coded 0. Census Bureau information on urban and rural places can be retrieved from the following internet site: http://www.census.gov/geo/www/tiger/index.html.

Users should note that this variable includes an "unknown" category. This value is assigned to respondents whose zip code includes both urban and rural areas or whose residence cannot be identified precisely enough to classify it as urban or rural. Respondents without valid address data are assigned a value of –3, invalid skip. Respondents who live out of the country are assigned a value of –4, valid skip.

COLLAPSED UNEMPLOYMENT RATE

Variable Created: UNEMPRATE-COL

To provide a measure of the economic situation in the respondent's area of residence, the dataset includes a variable indicating the unemployment rate. The round 1 NLSY97 unemployment rate variable was constructed using state and metropolitan area labor force data from the May 1998 publication of *Employment and Earnings* for the month of March 1998. The round 2 data were taken from the May 1999 publication for March 1999, the round 3 data were based on the June 2000 publication for March 2000, and the round 4 data were drawn from the June 2001 edition for March 2001. *Employment and Earnings*, published by the U.S. Department of Labor, Bureau of Labor Statistics, lists the size of the civilian labor force and number of unemployed persons for every state and most metropolitan areas. The variable is created as follows:

1. If the respondent lives in a metropolitan area that is listed in *Employment and Earnings*, then the unemployment rate in the NLSY97 variable is the unemployment rate for that metropolitan area. This rate is calculated by dividing the number of unemployed persons by the number of people in the civilian labor force as reported by BLS.

2. If the respondent does not reside in a metropolitan area listed in *Employment and Earnings*, he or she is assigned a "balance of state" unemployment rate. In these cases, the figures provided for the state and its metropolitan areas are used to compute the unemployment rate for the portion of the state that is not represented in any metropolitan statistical area. (Because the *Employment and Earnings* numbers are based on a different set of MSA codes than the NLSY97 geographic variables, there are a few cases in which NLSY97 metropolitan areas do not match those used in the BLS publication. Researchers who need more exact information should contact BLS or NLS User Services about completing a confidentiality agreement and obtaining the NLSY97 Geocode CD-ROM.)

After the MSA or balance-of-state unemployment rate is calculated for each respondent, the variable for the main file CD-ROM is collapsed into ranges (less than 3.0%, 3.0–5.9%, 6.0–8.9%, 9.0–11.9%, 12.0–14.9%, and 15.0% or higher). This collapsed variable protects the privacy and confidentiality of respondents.